FUZZY DECISION TREE-BASED INFERENCE SYSTEM FOR LIVER DISEASE DIAGNOSIS

Himaja Sivaraju

Columbus State University

TSYS School of Computer Science

The Graduate Program in Applied Computer Science

Fuzzy Decision Tree-based Inference System for Liver Disease Diagnosis

A Thesis in

Applied Computer Science

By Himaja Sivaraju

Submitted in Partial Fulfillment

Of the Requirements

For the Degree of Master of Science

December 2016

Abstract

Medical diagnosis can be challenging because of a number of factors. Uncertainty in the diagnosis process arises from inaccuracy in the measurement of patient attributes, missing attribute data and limitation in the medical expert's ability to define cause and effect relationships when there are multiple interrelated variables. Given this situation, a decision support system, which can help doctors come up with a more reliable diagnosis, can have a lot of potential.

Decision trees are used in data mining for classification and regression. They are simple to understand and interpret as they can be visualized. But, one of the disadvantages of decision tree algorithms is that they deal with only crisp or exact values for data. Fuzzy logic is described as logic that is used to describe and formalize fuzzy or inexact information and perform reasoning using such information. Although both decision trees and fuzzy rule-based systems have been used for medical diagnosis, there have been few attempts to use fuzzy decision trees in combination with fuzzy rules. This study explored the application of fuzzy logic to help diagnose liver diseases based on blood test results. In this project, inference systems aimed at classifying patient data using a fuzzy decision tree and a fuzzy rule-based system were designed and implemented. Fuzzy decision tree was used to generate rules that formed the rule-base for the diagnostic inference system.

Results from this study indicate that for the specific patient data set used in this experiment, the fuzzy decision tree-based inferencing out performed both the crisp decision tree and the fuzzy rule-based inferencing in classification accuracy.

Acknowledgements

I would like to express my sincere gratitude to Dr. Shamim Khan for his unwavering support and mentorship throughout my thesis. Dr. Khan shared his expertise with me very generously and I have learnt a lot from him. I would like to extend my thanks to Dr. Wayne Summers, Chair of TSYS School of Computer Science, and Dr. Rania Hodhod, Assistant Chair of TSYS School of Computer Science whose stimulating motivation and valuable ideas helped me to complete my thesis. Without the guidance of all of the above faculty members, completing my thesis would have been difficult. I am also grateful to the School of Computer Science for providing me a great environment in which I have constantly learned new things and grown as a better individual. I am also grateful to all faculty members who supported me on this journey. Last, but not the least, I would like to thank my parents for believing in me always and encouraging me to explore new horizons.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Data mining is the analysis of large amounts of data to discover relationships and patterns that have not been previously discovered by using tasks such as anomaly detection, association rule learning, clustering, classification, regression and summarization. Classification is one of the main tasks of data mining, which helps classify data into different meaningful categories based on a training set [6].Data mining techniques are widely used even in the medical field[12].

Death from cancer is increasing across the world [30], and one of the most common causes of cancer-related deaths is liver cancer [13]. Every year in America, almost 15,000 people die from liver disease [1]. In underdeveloped countries, resources to detect liver cancer are limited [31], misdiagnosis and lack of availability of health professionals are some other factors which contributes to the delay of early diagnosis of cancer. Late diagnosis can be one of the primary reasons for this increase in the death rate due to cancer[33].An early diagnosis of the cancer may increase patient's survival rate [32].Given this situation, an automatic tool to diagnose any liver disease based on blood test results could be helpful.

Data mining techniques are one of the broadly used techniques in building decision support systems, especially clinical decision support systems [11].A decision based system to diagnose liver diseases could help doctors with early and more accurate diagnosis, thus decreasing the chances of misdiagnosis. It should be simple such that even general practitioners are able to use it to give advice to patients about the urgency of consulting a specialist if needed.

1.1 Fuzzy Logic

The idea of fuzzy logic was brought to light by Dr. Lotfi Zadeh of the University of California in the 1960s, while he was working on the problem of computer understanding of natural language [14]. Fuzzy logic is described as logic that is used to describe and formalize fuzzy or inexact information. It is an approach to computing based on degrees of truth rather than the true or false values expected in classical Boolean logic. Instead of interpreting facts as absolutely true or absolutely false, fuzzy logic accepts that they can be partly true and partly false at the same time. Application of fuzzy logic has proved itself to be a powerful technique for decision making in many areas [15]. One such area where fuzzy logic has been found useful is medical diagnosis [16].

In general, variations in diagnostic decisions made by medical practitioners arise because of uncertainties or vagueness in patient information used in the diagnostic process. In general, practitioners consider cause and effect relations that tend to give poor results because of its inability to deal with uncertainty in data. In such cases, a fuzzy expert system can be useful.

Using techniques such as fuzzy logic can be used in medical diagnosis as more reliable conclusions can be made relating to individual patient's data. Fuzzy logic can be a powerful technique especially in liver, heart, and diabetes disease diagnosis [8] [16] [17] [18]. A medical diagnosis in such diseases can be a complicated task and needs to be performed more reliably. Fuzzy logic in liver disease diagnosis can help capture the required medical information of patient and come up with diagnosis decisions that are more accurate than traditional approaches in the medical world [18]. Practically in today's medical world, an expert physician's experience is specified in fuzzy terms, which helps them to portray the accurate knowledge rather than knowledge with uncertainties [10]. Generally, experts tend to use fuzzy terms while interacting

with the patient. For example, if a patient visits the doctor's office. The doctor may use fuzzy terms like "your fever is very high" instead of saying "your fever is 103 Celsius".

1.2 Decision Trees

Decision trees (DTs) are tree-like structures; they are used in data mining for classification and regression [6]. The goal of DTs is to create a model, which can help predict a target variable based on several input variables. After developing the model, rules can be derived following the path from the root node to the corresponding leaf node; conjunction or disjunction can be used to connect the internal nodes. DTs are simple to understand and interpret as they can be visualized. But, one of the disadvantages of this algorithm is it considers only crisp values for both input and target class.

Figure 1 is a sample decision tree, which was generated using a hepatitis liver dataset by Shankar Sowmien [5].The attributes such as *age*, *sex*, *liver big*, *liver firm*, *albumin*, *SGOT*, *spiders* are used to build the decision tree where each attribute has its own range. For example, age attribute has a range from 10 to 90. The classification of this tree is based on the all patient attributes and the class, which helps in classifying whether the patient lives or dies after the diagnosis: One such rule generated by the decision tree is

If Ascites <=1 and Albumin > 2.8 and liver firm <=1 then patient lives.

Figure 1.Example of a crisp decision tree

In order to build any decision tree using ID3 algorithm, Decision tree is usually built from the root node by dividing into the subsets with homogeneous instances and to calculate this homogeneousity of a sample, entropy used. Entropy of each branch is calculated and the branch which has entropy value as zero is considered as root node. After the entropy is calculated, information gain is calculated which is decrease in entropy. The attribute with largest information gain is considered as next decision node and this selection of next attribute continues until all the data is classified.

1.3 Fuzzy Decision Trees

Fuzzy decision trees combine the crisp decision tree with fuzzy set theory, i.e. the crisp values used to split the decision criteria are replaced with fuzzy values. To make rules from the decision trees, the path from the root node to the corresponding leaf node is considered. Conjunction or disjunction is normally used to connect the internal nodes. The same methodology is applied to decision trees using fuzzy logic concepts. The decision on splitting a branch in fuzzy decision

trees can be based on either numerical or linguistic values. This is one of the main advantages of fuzzy decision trees [3].

Rules derived from fuzzy decision trees are easy to read and understand. Each node in a fuzzy decision tree is associated with a variable; each branch is associated with a fuzzy subset of that variable. Hence, rules derived from fuzzy decision trees are fuzzy rules. Let's consider an example rule from the crisp decision tree shown in Figure 1 above.

If Ascites <=1 and Albumin <= 2.8 Then Patient Dies

The above rule requires specific threshold values that can be difficult to derive or agree on. Fuzzy rules accept inexact linguistic values in rules. Using fuzzy rules instead of rules from crisp decision tree can make it a lot easier to understand the rules as they are expressed in linguistic terms such as high, low, medium etc., instead of crisp numerical values. Given below is an example of a fuzzy rule derived from fuzzy decision trees [8]:

If Age=" Young" ANDVR_HG=" High" Then Class=" High"

The goal of this thesis is to develop an inference system to aid the diagnosis of liver diseases using fuzzy logic where rules are derived from fuzzy decision trees. Unlike traditional approaches, which give uncertain outputs, we propose to use fuzzy logic to come up with a more accurate diagnosis of the patient's disease. Here fuzzy logic will be used to generate a fuzzy inference system (FIS), which uses fuzzy set theory to map inputs to outputs. The output of the system will be the probability of a patient having liver disease and inputs will be various test results related to liver diagnosis such as Total Bilirubin, SGOT aspartate aminotransferase, SGPT alamine amino transferase, Albumin, and andalkphos alkaline phosphatase.

1.4 Fuzzy Inference Systems

Fuzzy logic can be used to generate a fuzzy inference system (FIS), which uses fuzzy set theory to map inputs to outputs. In case of medical diagnoses, inputs can be patients' test results, and the output is the diagnostic result. In order to compute the output of the FIS, given the inputs, one must go through the following steps:

1. Inputs are fuzzified using the input membership functions. Membership function is used to graphically represent the input points, and it helps to describe how each input point is mapped to a membership value. Fuzzification is the first and foremost step to build an inference system It is concerned with transforming the input to fuzzy sets with the help of membership functions is called fuzzification.
2. A rule base must be built. A rule base is composed of if - then rules. These rules help in transforming inputs to output, i.e. based on a rule and inputs, the diagnosis of the patient is determined.
3. If a crisp output is needed, then the output should be defuzzified. Defuzzification is process of transforming fuzzy sets into crisp values. Centroid, bisector, middle, and smallest and largest of maximum are examples of defuzzification methods.

1.5 Thesis Goals

These days, it is a real challenge for a physician to go through patient reports and diagnose the disease due to time constraints. The expert needs to take into account numerous symptoms and diagnostic measurements.They also have to deal with complex relationships between multiple interacting factors; uncertainty introduced by errors in measurement, and missing data. In this situation, a decision support system which can help in assisting an expert to diagnose the disease can come in handy. For instance, this system can help the expert in recommending a preliminary

diagnosis which can be used like a second opinion. It can even be of help to the certified general physician in advising the patient if he/she should consult the specialist or not.

This study aims to validate the following hypotheses:

1. Fuzzy decision trees can outperform crisp decision tree-based systems in the accuracy of classification applied to liver disease diagnosis.

2. Fuzzy decision trees can provide the necessary rules to build a fuzzy rule-based decision support system to diagnose liver disease.

3. Fuzzy rule-based inference systems can outperform fuzzy decision trees in accuracy of classification applied to liver disease diagnosis.

1.6 Thesis Organization

This thesis focuses on liver diagnosis so a literature review on the use of decision trees and fuzzy inference systems will be covered in chapter 2. Chapter 3 discusses the methodology used and how the proposed system is implemented. Chapter 4 covers the experimental results. Finally, conclusion and future work are presented in chapter 5.

Chapter 2.Related Work

Different researchers used different methodologies in the field of medical diagnosis. Some of the applications which are developed so far using decision trees and fuzzy inference systems are discussed below:

Liver disease can be assessed by using various algorithms like the linear discriminant analysis (LDA), diagonal linear discriminant analysis (DLDA), quadratic discriminant analysis (QDA), diagonal quadratic discriminant analysis (DQDA), naive Bayes (NB), feed-forward neural network (FFNN), and classification and regression tree (CART). These Algorithms help to identify the exact problem associated with the liver and guides the doctors for better treatment based on the results generated from the assessment algorithms. Of all the above algorithms, results obtained by CART are proven to be much more accurate, reducing the inefficiency in the results. The accuracy of the results mostly depends on the type of datasets used as input for the algorithms [35].

Different techniques were used for different purposes as explained below.

Artificial Neural Networks is used in the field of medical diagnosis as follows:

- For the early identification of hepatectomised patient [36].
- To cure the hepatobiliary disorders [37].
- For the hepatitis disease diagnosis [38].
- For the classification of liver cyst, hepatoma and cavernous haemangioma [39].
- To diagnose types of cirrhosis [40].
- For the classification of fatty liver, liver cirrhosis and liver cancer [41].

Artificial Immune System is used for the diagnosis of Hepatitis disease. ANN-CBR together is used for knowing the types of liver disorder and their treatment [40].

Fuzzy logic was used to classify liver disorders under Alcoholic liver damage, primary hepatoma, liver cirrhosis and cholelithiasis, to differentiate between healthy and unhealthy liver patients, to diagnose hepatitis, and to perform semi-automatic liver tumor segmentation [47].After careful examination of all the techniques, It was observed that novice researchers use methodologies such as ANN and Artificial Neural Network combined with fuzzy logic for liver disorder datasets as it has wide acceptance with higher accuracy results [47].

Sow mien et al. proposed a diagnosis system for a type of liver disease called hepatitis using machine learning. They used the C4.5 algorithm to generate a decision tree to find the abnormalities of patient with 19 attributes and obtained an accuracy of 85.81 with their overall study [5].

Kumar and Sahoo wrote a paper in which they used Support Vector Machines (SVM), rule induction, decision trees, Naive Bayes classification, Artificial Neural Networks (ANN), and data mining with K-cross fold techniques for the prediction of liver diseases. The results from their experiments showed that a rule-based classification model with decision tree techniques gave most precise results [9].

In the paper, "Improving the Prediction Accuracy of Liver Disorder Disease with Oversampling," Hyontai experimented on 'BUPA liver disorder' dataset with C4.5 and CART decision trees algorithm. They validated the over-sampling method in minor classes to effectively deal with the data insufficiency problem. The results proved that the oversampling method is effective, and it is more effective when used in the CART decision tree algorithm [4].

Parminder and Aditya used random tree algorithm to generate rules and decision trees for the classification of liver based diseases. Based on the attributes like Neurological, Psychiatric, Pathological, Physical and cognitive and symptoms, the authors generated decision tree using Weka to diagnosis the liver based diseases, such as Wilson, fatty liver, Cholesteric, inherited, and autoimmune. The results show that decision trees can be used to model actual diagnosis of liver cancer for surgical and non-surgical treatment [49].

Eyke Hullermeier performed a survey on why Fuzzy Decision Trees are Good Rankers. The author discovered that Laplace correction significantly increases performance in terms of AUC and un-pruned trees almost always have higher AUC values than standard pruned trees (Laplace correction). In other words, according to the author a single decision tree cannot be both a good classifier and a good ranker at the same time. Whereas, author thinks that Fuzzy decision trees may overcome this problem [50].

Some studies were carried out on medical diagnosis using decision trees and fuzzy logic. Decisions trees were used for cardiovascular dysautonomias diagnosis [8]. The researchers developed a fuzzy decision tree based on patient dataset and compared error rates between crisp and fuzzy decision trees. Their results showed that fuzzy decision trees were better in accuracy than crisp decision trees. Fuzzy decision trees were also used for prediction of the death of a patient with heart failures and the experimental results of this research were shown to be accurate with a sensitivity of 67.3% and a specificity of 62.6% [3].

Among the different techniques used to diagnose the liver disorders, using fuzzy decision trees is preferred for the following reasons [48]:

1. Like the other techniques fuzzy decision trees not only work on true/false values they also work on uncertainty values. For example, if the data values are in between these ranges consider the disease to be like hepatitis or an infection.
2. FL systems are reliable, easy to understand, analyze and train.
3. They can work even with the imprecise and ambiguous data.
4. They work with global-K and fast global-K that gives them the feature of even working with the datasets that high noise and inaccuracy in them.

From the Literature review, it was shown that fuzzy decision trees has many advantages such as it can perform better in accuracy and a good classifier compared to crisp decision tree. Moreover, it was found that artificial neural networks combined with fuzzy logic can be more appropriate to diagnose liver disorders [47].

Chapter 3. Methodology and implementation

3.1 Methodology

In order to develop a system designed to perform preliminary diagnosis of the liver disease, three separate inference systems for classification were built: a crisp decision tree, a fuzzy decision tree, and a hybrid system that uses a fuzzy decision tree to derive a set of rules that will subsequently be used in a fuzzy rule-based inference system. All three systems were then evaluated and compared in terms of their accuracy and ease of understanding. The steps given below were followed to build the three inference systems:

i. Data Acquisition and pre-processing.

ii. Construction and evaluation a crisp decision tree-based inference system.

iii. Construction and evaluation of a fuzzy decision tree-based inference system.

iv. Construction and evaluation of the hybrid fuzzy rule-based inference system.

3.2 Implementation

Step 1: Data Acquisition and Pre- processing

A public dataset, ILPD (Indian Liver Patient Dataset), consisting of 583 patient records, was downloaded from the UCI Machine Learning Repository [19]. The dataset was collected from the Northeast of Andhra Pradesh, India, and contains 416 records of patients diagnosed with liver disease and 167 records of patients diagnosed to be free from liver disease.

	A	B	C	D	E	F	G	H	I	J	K	L
1	AGE	GENDER	TB	DB	AAP	SGPT	SGOT	TP	ALB	A/G	CLASS	
2	66	Male	0.6	0.2	100	17	148	5	3.3	1.9	1	
3	27	Male	1	0.3	180	56	111	6.8	3.9	1.85	1	
4	63	Male	0.9	0.2	194	52	45	6	3.9	1.85	1	
5	28	Female	0.9	0.2	316	25	23	8.5	5.5	1.8	0	
6	35	Female	0.9	0.2	190	40	35	7.3	4.7	1.8	1	
7	43	Female	0.9	0.3	140	12	29	7.4	3.5	1.8	0	
8	62	Male	5	2.1	103	18	40	5	2.1	1.72	0	
9	48	Male	0.7	0.2	326	29	17	8.7	5.5	1.7	0	
10	17	Female	0.5	0.1	206	28	21	7.1	4.5	1.7	1	
11	50	Male	1.1	0.3	175	20	19	7.1	4.5	1.7	1	
12	25	Female	0.9	0.3	159	24	25	6.9	4.4	1.7	1	
13	50	Male	0.9	0.2	202	20	26	7.2	4.5	1.66	0	
14	40	Female	2.1	1	768	74	141	7.8	4.9	1.6	0	
15	68	Male	1.8	0.5	151	18	22	6.5	4	1.6	0	
16	54	Male	2.2	1.2	195	55	95	6	3.7	1.6	0	
17	31	Male	0.6	0.1	175	48	34	6	3.7	1.6	0	
18	31	Male	0.6	0.1	175	48	34	6	3.7	1.6	0	
19	29	Male	0.7	0.2	165	55	87	7.5	4.6	1.58	0	
20	70	Male	1.4	0.6	146	12	24	6.2	3.8	1.58	1	
21	17	Male	0.9	0.2	224	36	45	6.9	4.2	1.55	0	
22	28	Female	0.8	0.2	309	55	23	6.8	4.1	1.51	0	
23	37	Male	0.8	0.2	195	60	40	8.2	5	1.5	1	
24	24	Male	1	0.2	189	52	31	8	4.8	1.5	0	

Figure 2. A snapshot of Dataset

The dataset contains 10 attributes: age of the patient, gender, total bilirubin, direct bilirubin alkaline phosphotase, alamine aminotransferase, aspartate aminotransferase, total proteins, albumin, and albumin and globulin ratio. Each record in this data was already classified by experts and the diagnosis is stored in a variable called *class label*. The value for each attribute is a binary number; 0or1, where 0 indicates that the patient is a liver patient and 1 indicates that the patient is not a liver patient. Each attribute in the dataset has its own significance in diagnosing the liver disease as described in the table 1 below:

Table 1. Attributes and their significance in diagnosing liver disease

Attribute	Significance
Total Bilirubin	Total Bilirubin test is used to detect the levels of bilirubin in the blood i.e. either increased or decreased in the blood. This test helps to identify the presence of jaundice in the blood i.e., too much Bilirubin in the blood can cause jaundice or icterus and this test also helps in identifying the various liver syndromes [20]. Generally, bilirubin is a precipitate produced by heme, which in turn is produced by hemoglobin of Red Blood Cells (RBC). This Bilirubin is filtered by liver and in the cases where liver cannot filter this wastage or when excess amounts of it is produced then it leads to the malfunction of the liver causing liver related diseases[20]. Bilirubin is classified into two types. 1. Unconjugated Bilirubin (it is produced by hemoglobin and carried to the liver through proteins) and 2. Conjugated Bilirubin (usually not present in the blood, these are produced in the liver when sugars are attached to Bilirubin) [20].
Serum glutamic oxaloacetic transaminase (SGOT) *Aspartate Aminotransferase*:	SGOT test plays an important role in knowing enzyme levels in the blood. SGOT enzyme is present in RBC, liver, heart, pancreas and various other major organs. Levels of Aspartate Aminotransferase (AST) increases in the blood when an organ gets damaged. The more the extent of damage of organ, the more AST that is released into the blood. This test is usually done to monitor the patient with liver disease [21].
SGPT (Serum glutamic pyruvic	SGPT test is one of the most important test and more specific to liver than any other tests. This enzyme is mostly present in liver with minor

transaminase) Alamine Aminotransferase	amounts in kidneys, pancreas etc. Usually low levels of alamine aminotransferase are observed in blood. But when liver or other parts containing alanine aminotransferase are damaged, then more alanine amino transferase are released into the blood. Both the tests of SGPT and SGOT can be done at same time and the ratio of SGOT and SGPT helps to find out whether the liver is damaged or not [22].
Albumin	Albumin test is used to evaluate the function of liver and kidneys i.e., albumin test can help in knowing if the body is absorbing sufficient proteins or not. It plays an important role in preventing fluid in the blood leaking into the tissues. Lower levels of Albumin are a warning that further analysis might be required. Dehydration and high protein diet increase the albumin levels in the blood [23].
Alkphos *Alkaline Phosphatase (AAP)*:	AAP test is used to measure the amount of alkphos enzyme in the blood and helps to determine how well the liver is functioning and also for the bone disorders. Usually the blood has low levels of ALP, but when the person has liver or bone disorders then more ALP is observed in that person's blood. Checking the alkphos levels is necessary for a liver function test and helps to determine if liver is damaged or diseased [24].
Total Proteins	Total Proteins test is used for the calculation of total proteins and globulins present in the body. Generally, this test is performed on the patients who have weight loss, kidney and liver diseases. Less number of proteins in the body leads to fatigue, kidney disorders and liver malfunctions [25].

Direct Bilirubin Alkaline Phosphotase:	Direct bilirubin test is used to check malfunction of the liver. Usually the blood has lower levels of alkaline phosphatase. Increased levels of malfunctioning of the liver leads to increased levels of alkaline phosphatase in the blood. This test is always performed in conjunction with Aspartate Aminotransferase and Alanine Aminotransferase.
Albumin and Globulin Ratio (A/G ratio):	A/G ratio test is most of the times performed in conjunction with the total proteins test. This test is used to analyze the albumin and globulin ratio If A/G ratio is not in the normal range then there are the chances of patient going through fatigue and bone disorders [26].

Data pre-processing helps convert the raw data into system acceptable format. Input data needs to be processed before feeding it into the system; this pre-processing includes data cleaning, normalization, transformation, feature extraction …etc. To help with this research, data cleaning was performed. Records which had missing and abnormal values were removed. The general process involved two phases as described below:

a. Dealing with missing data:

Not all the times the missing data has to be considered for the experimental process. It all depends on the impact of missing data on the final response. If the missing data has nothing to do with the final output then it can be neglected and if the same missing data plays a crucial role in the final output then it has to be given utmost importance.

This problem can be overcome by two methods [27].

1. Deleting the missing data: If the missing data doesn't have any impact on final output or if the data is missing at random intervals then such data can be removed.

2. Filling the missing data based on the rest of the datasets: In this case the average of the available data sets is taken and replace those with the missing values.

It was found that there were only four data records with missing values out of 583 data records in total. As this is a low number compared with the total number of records, those four data records were just deleted.

b. Dealing with outliers:

An outlier can be considered as an odd value or a distracting value in the dataset that reduces the analytical capability of a dataset. One of the ways to deal with the outliers is to delete them.

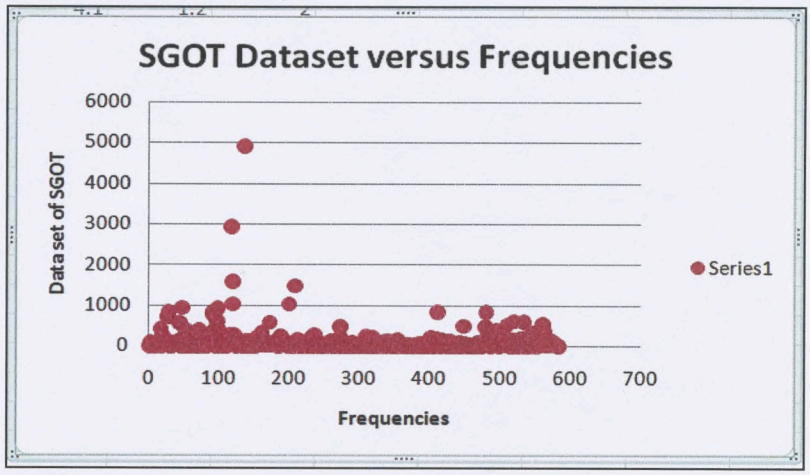


Figure 3. SGOT dataset versus frequencies

Figure 3 shows a visualization of one of the attributes called SGOT in the dataset that is used in this research. The outliers are observed at two points after the data visualization. One at 4929 and the other at 2946 and the rest of the values are under the value of 2000. Hence, these two outliers were deleted from the data.

Step 2: Construction and evaluation of Crisp Decision Tree

As shown earlier in Figure 2, there are 10 input attributes and one selector field labeled by experts. Each record in the patient dataset is already classified based on the corresponding patient's diagnosis by medical experts. This classification is represented by the column headed "Class", with a 0 label indicating that the patient is a liver patient and 1 indicating that the patient is not a liver patient. The decision tree helped in deriving the combinations of input attributes by which the class was labeled.

By trial and error method, crisp decision trees were developed using Weka 3 which is free-ware software available online used to build the crisp decision tree using the Indian Liver Patient dataset [29]. Two crisp decision trees were developed one tree with fewer rules and lower accuracy and other tree had more rules and higher accuracy. Both the decision trees were built using C 4.5 algorithm. At each and every node of the tree, C4.5 chooses the attribute which has the highest normalized information gain to make the decision. A visualization of the crisp trees is shown in Figure 4 and Figure 5. The decision tree shown in Figure 4, has 19 rules with 69% accuracy, while the decision tree shown in Figure 5 has 1547 rules with accuracy equals 86.44%.

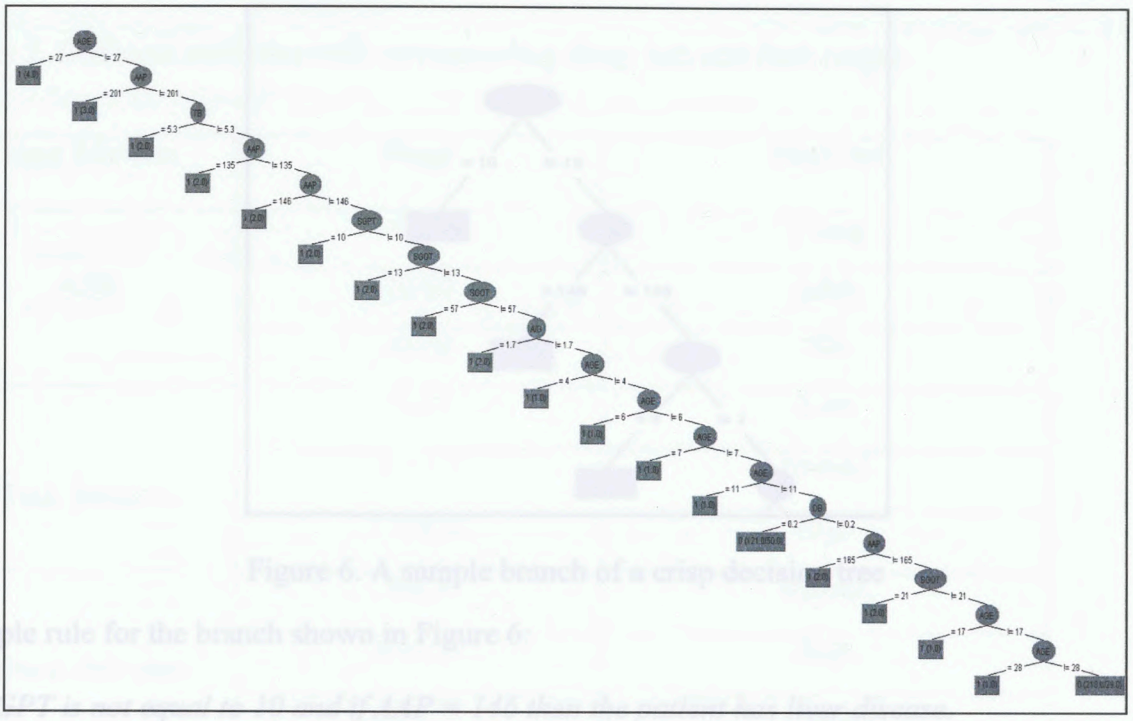


Figure 4. The crisp decision tree with 19 rules

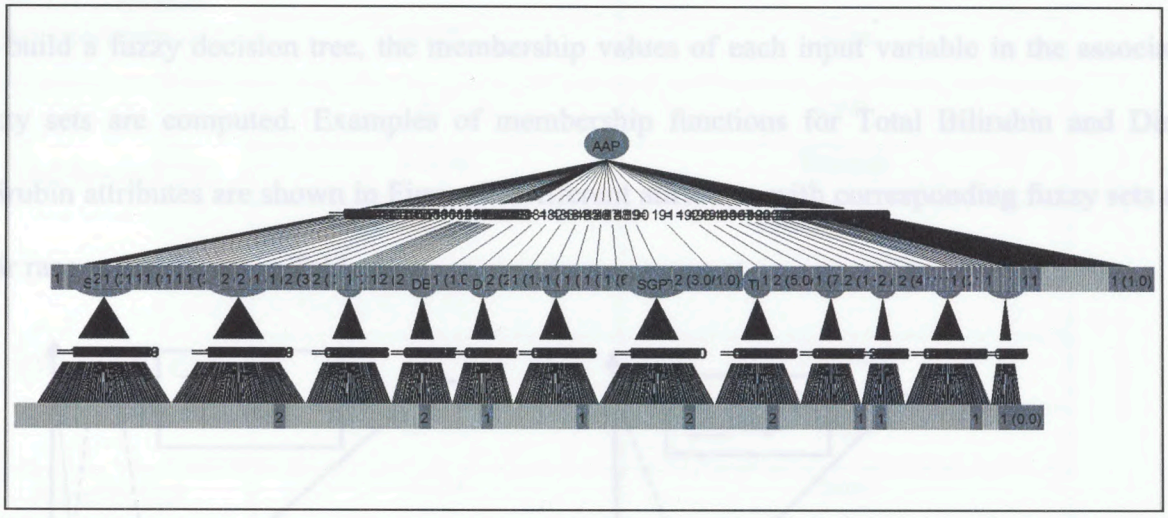


Figure 5. Crisp decision tree with 1547 rules

To derive the rules from a crisp decision tree, each path from root node to leaf node is considered. An example rule can be attained by following a path from the root node to a leaf node, which represents a classification as can be seen in Figure 6

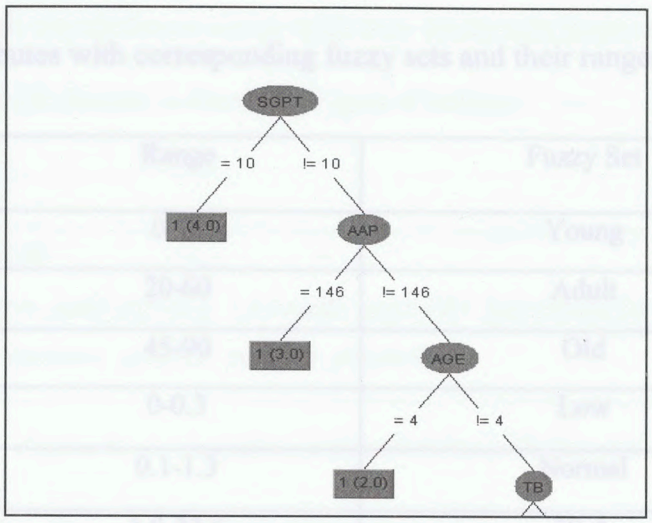


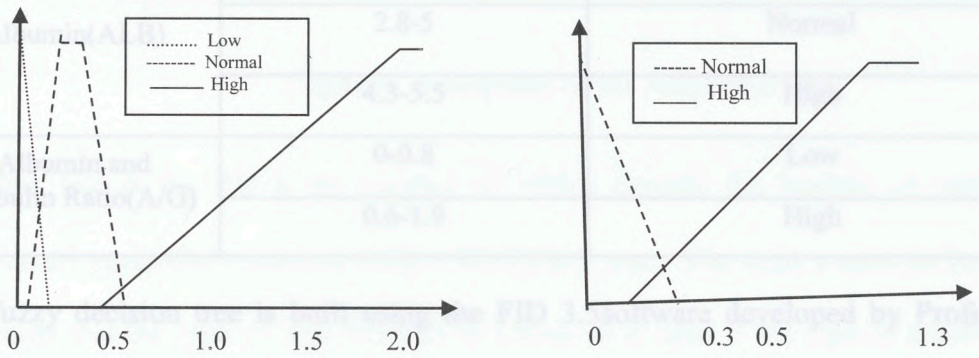
Figure 6. A sample branch of a crisp decision tree

Sample rule for the branch shown in Figure 6:

If SGPT is not equal to 10 and if AAP = 146 then the patient has liver disease.

Step 3: Construction and evaluation of the fuzzy decision tree

To build a fuzzy decision tree, the membership values of each input variable in the associated fuzzy sets are computed. Examples of membership functions for Total Bilirubin and Direct Bilirubin attributes are shown in Figure 7. Different attributes with corresponding fuzzy sets and their ranges are shown in Table 2.



(a) Membership for the attribute Total Bilirubin (b) membership for the attribute Direct Bilirubin

Figure 7. Membership functions for Total and Direct Bilirubin

Table 2. Different attributes with corresponding fuzzy sets and their ranges

Input Attribute	Range	Fuzzy Set
AGE	0-25	Young
	20-60	Adult
	45-90	Old
Total Bilirubin	0-0.3	Low
	0.1-1.3	Normal
	1.0-32.6	High
Direct Bilirubin	0-0.3	Normal
	0.2-1.3	High
	30-875	High
SGOT Aspartate Aminotransferase	0-40	Normal
	30-950	High
Total Protein(TP)	0-7	Low
	5-8.35	Normal
	7.9-9.2	High
Albumin(ALB)	0-3.2	Low
	2.8-5	Normal
	4.3-5.5	High
Albumin and Globulin Ratio(A/G)	0-0.8	Low
	0.6-1.9	High

The fuzzy decision tree is built using the FID 3.5software developed by Professor Cezary Z. Janikow [33]. This software requires three files to build the fuzzy decision tree:

1) Attributes file, 2) Events file, and 3) Templates file.

The first file contains all the attributes along with their fuzzy sets were converted into a specific format. Snapshot of the file format is shown in Figure 8 below;

```
NumberOfAttributes
AttrName attrType numLingVals [lowerBd upperBd [minNumVals maxNumVals]]
   LingValName [point1 point2 point3 point4]
```

Figure 8. A snapshot of the attributes file format

```
File  Edit  Format  View  Help
10

AGE 1 3 0 90
Young 0 0 0.1 0.3
Adult 0.2 0.3 0.4 0.6
Old 0.5 0.9 1 1

GENDER 0 2
Male
Female

TB 1 3 0 32.6
Low 0 0 0.01 0.012
Normal 0.01 0.02 0.03 0.043
High 0.03 0.5 1 1
```

Figure 9. A snapshot of the Attributes file

The first line in the file is the number 10 which denotes the number of attributes. The file contains all attributes names as AGE, GENDER …etc. The digit 1 next to the AGE attribute represents that the type of Age attribute is Linear (0 is used if the type of attribute is nominal). The digit 3denotes that there are three linguistic values (represented by three fuzzy sets) associated with the AGE attribute. The range 0 - 90 represents range of values the Age attribute can take. Young, Adult and Old are the names of fuzzy sets associated with the attribute Age.

In the second file, the entire data set is described. 450 records (referred to events in FID terminology) were used to train the tree and 105 events to test the tree.

Snapshot of the event file is shown below:

```
NumEvents numOfAttributes
   Attr1Value Attr2Value ... decisionValue weightValue
   ...
```

Figure 10. A snapshot of the Events file

Variable Decision-Value represents the classification of each patient data record into either one of two categories – diagnosed as a liver disease patient and not a liver disease patient.

A sample extract from the event file that was used in building the tree is shown below:

```
450     10

66      Male    0.6     0.2     100     17      148     5       3.3     1.9     1       1
27      Male    1       0.3     180     56      111     6.8     3.9     1.85    1       1
63      Male    0.9     0.2     194     52      45      6       3.9     1.85    1       1
28      Female  0.9     0.2     316     25      23      8.5     5.5     1.8     0       1
35      Female  0.9     0.2     190     40      35      7.3     4.7     1.8     1       1
43      Female  0.9     0.3     140     12      29      7.4     3.5     1.8     0       1
62      Male    5       2.1     103     18      40      5       2.1     1.72    0       1
48      Male    0.7     0.2     326     29      17      8.7     5.5     1.7     0       1
17      Female  0.5     0.1     206     28      21      7.1     4.5     1.7     1       1
50      Male    1.1     0.3     175     20      19      7.1     4.5     1.7     1       1
25      Female  0.9     0.3     159     24      25      6.9     4.4     1.7     1       1
50      Male    0.9     0.2     202     20      26      7.2     4.5     1.66    0       1
40      Female  2.1     1       768     74      141     7.8     4.9     1.6     0       1
68      Male    1.8     0.5     151     18      22      6.5     4       1.6     0       1
54      Male    2.2     1.2     195     55      95      6       3.7     1.6     0       1
31      Male    0.6     0.1     175     48      34      6       3.7     1.6     0       1
31      Male    0.6     0.1     175     48      34      6       3.7     1.6     0       1
29      Male    0.7     0.2     165     55      87      7.5     4.6     1.58    0       1
```

Figure 11.A sample output from the Events file

It is worth noting that as it was not known which event is more important; all the events were given equal weights.

The third file consists of the default of the template that can be used, all the parameters which help in building the tree like Chi-squared test, fuzzy stop level, type of discretization, etc., are defined here.

FID35 software allows both top down and bottom up discretization, top down approach was used in this research. Top-down discretization performs data-driven discretization by splitting sets spanned over linear non pre-partitioned attributes. Splitting is data-driven, i.e. it performs splitting only on relevant attributes. Set-based inferences were used in building the tree as the set based inferences treat leaves as fuzzy sets. Chi-squared test was performed.

Step 4: Construction and evaluation of the hybrid fuzzy inference system

To build the fuzzy inference system we need input data and rules. Input data was available from the Step 1. Rules play a very important role in building the FIS. Fuzzy rules extracted from Step 3 were used in building the fuzzy inference system. An example of how the following fuzzy rule was extracted is discussed below.

[DB=Normal][SGOT=Normal][TP=Normal][AGE=Adult][ALB=Normal][GENDER=Male][SGPT=Normal][TB=Normal][AAP=Normal]: IN=0.97 PN=1.96: Yes=0.78 No=1.18 RS=0.611 rs=1.000 bestDec=1

The set based inference system is an inference which uses the areas and/or centroids of the fuzzy sets to compute the decision value for the dataset. Here, in this paper the fuzzy decision tree uses the centroids of the fuzzy sets to compute the decision value. There are different stopping criteria's used to avoid producing large fuzzy decision trees, one among those is by using "IN" and "PN" in Figure 12.

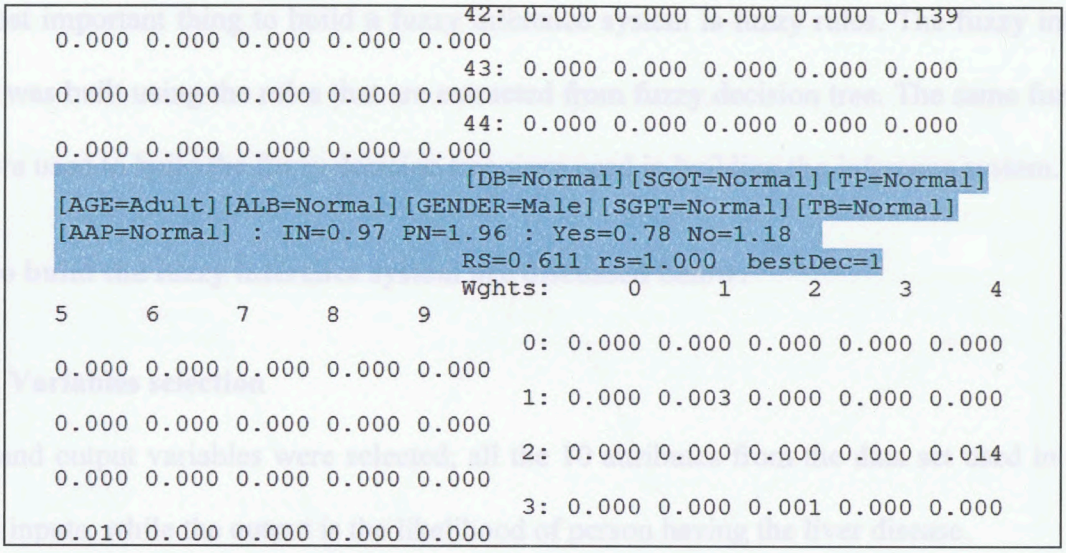


Figure 12. A sample screen shot of the Tree file generated by the FID software

IN helps in providing information content at which expansion should be stopped and PN helps in providing the minimal event count at which expansion should be stopped. Yes and No in Figure 12 above are the names of decision class, each of these decision classis followed by the centroid values, e.g., the shaded event in Figure 12has a centroid of 0.78 in the 'Yes' class and 1.18 in the 'No' decision. Best Dec is the delta best value; this uses centroid from majority class in leaf value. In this case it is '1' (having the highest centroid of 1.18) i.e., it belongs to 'No' decision class, as 0 represents decision class 'Yes' and 1 represents the decision class 'No'. From the above record of the tree file, we can derive the fuzzy rule as:

If [DB=Normal] and [SGOT=Normal] and [TP=Normal] and [AGE=Adult]and [ALB=Normal] and [GENDER=Male]and [SGPT=Normal]and [TB=Normal] and [AAP=Normal] then the probability of patient having liver disease is low

The most important thing to build a fuzzy inference system is fuzzy rules. The fuzzy inference system was built using the rules that are extracted from fuzzy decision tree. The same fuzzy sets that were used to build the fuzzy decision tree were used in building the inference system.

Steps to build the fuzzy inference system are discussed below:

Step 1. Variables selection

Inputs and output variables were selected; all the 10 attributes from the data set used in 2 were used as inputs, while the output is the likelihood of person having the liver disease.

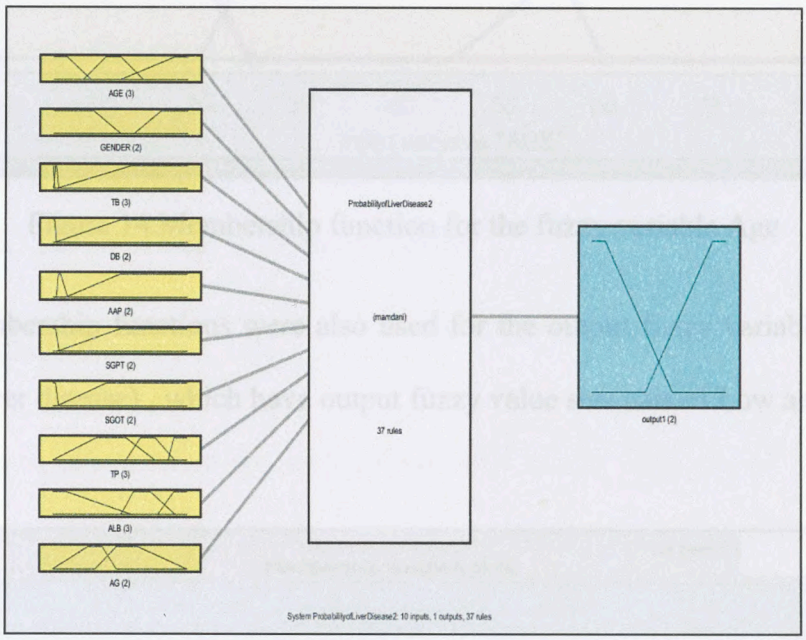


Figure 13. A snapshot of the fuzzy inference system

Step 2. Fuzzification

Fuzzifying the inputs is the first step to build any fuzzy inference system, this is done by transforming the input to fuzzy sets with the help of membership functions. Membership function is calculated for each fuzzy set, i.e. range values are estimated and the shape of the

function is chosen. In this FIS, Trapezoidal membership functions were used for the input variable Age. Age has three fuzzy set values: Young, Adult, and Old as shown in Figure 14.

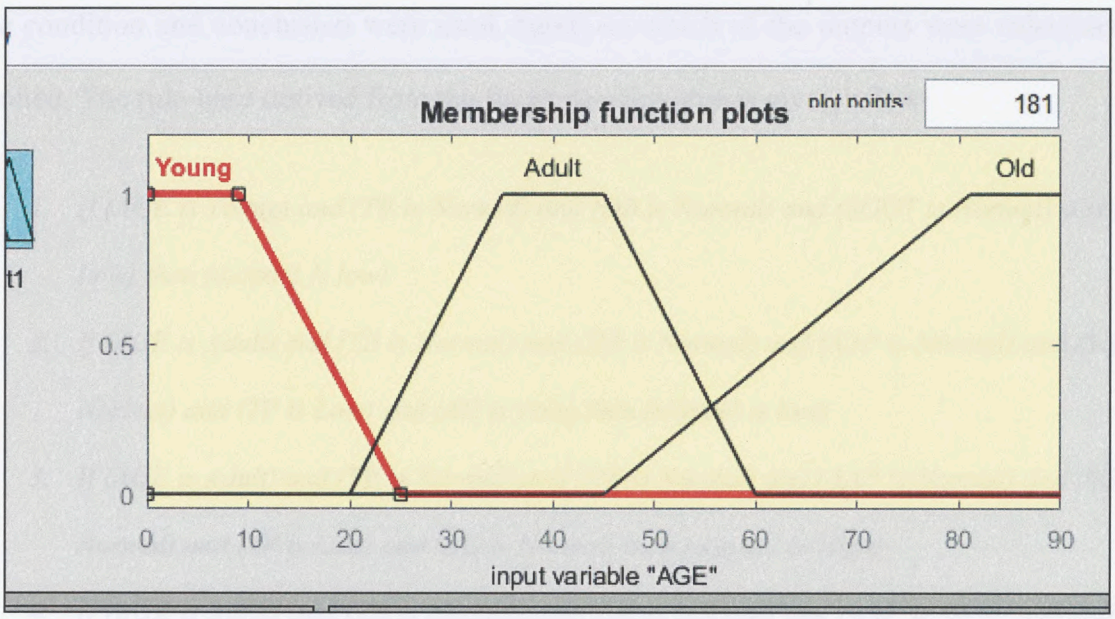


Figure 14.Membership function for the fuzzy variable Age

Trapezoidal membership functions were also used for the output fuzzy variable (Probability of patient having liver disease) , which have output fuzzy value sets named Low and High as shown in Figure 15

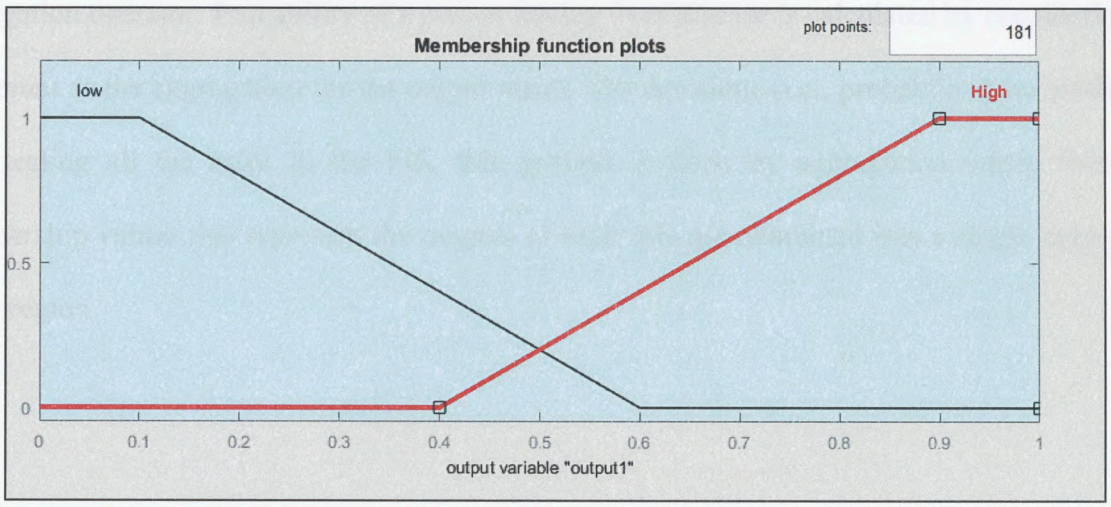


Figure 15. Membership function for the fuzzy variable Probability of liver disease

Step 3. Rule Evaluation

The rule-base was constructed from the rules extracted from the fuzzy decision tree.

If-else condition and conclusion were used, based on which of the outputs were calculated and controlled. The rule-base derived from the fuzzy decision tree is given below:

1. *If (AGE is Young) and (TB is Normal) and (DB is Normal) and (SGOT is Normal) and (TP is Low) then (output1 is low)*
2. *If (AGE is Adult) and (TB is Normal) and (DB is Normal) and (AAP is Normal) and (SGOT is Normal) and (TP is Low) and (AG is Low) then (output1 is low)*
3. *If (AGE is Adult) and (TB is Normal) and (DB is Normal) and (AAP is Normal) and (SGOT is Normal) and (TP is Low) and (AG is Normal) then (output1 is High)*
4. *If (AGE is Adult) and (TB is Normal) and (DB is Normal) and (AAP is High) and (SGOT is Normal) and (TP is Low) then (output1 is low)*

Step 4: Defuzzification

In this step the outputs obtained for each rule in step into a single fuzzy set, using a fuzzy aggregation operator. Probability of a person having liver disease is calculated by considering the maximum as the aggregation for the output result. The decisions (i.e., probability) are made only after testing all the rules in the FIS, this process is done by aggregation where fuzzy set membership values that represent the outputs of each rule are combined into a single aggregated fuzzy region.

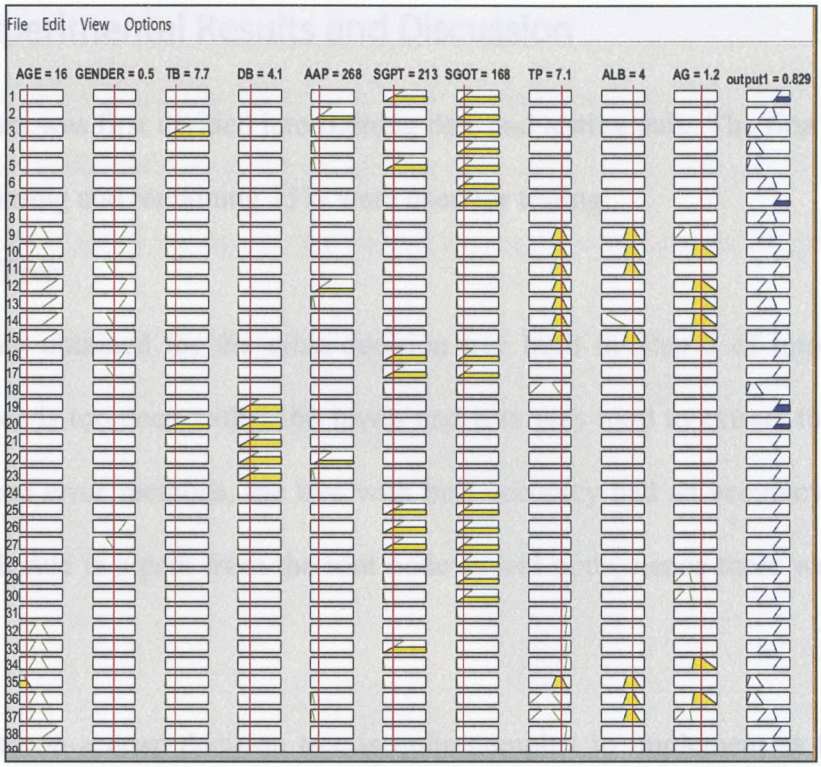


Figure 16. Output of Fuzzy Inference System

The patient data described in Table 3 below is the sample records from the patient dataset used in this research.

Table 3. Sample patient record

AGE	GENDER	TB	DB	AAP	SGPT	SGOT	TP	ALB	A/G	System Output
16	Male	7.7	4.1	268	213	168	7.1	4	1.2	82.9%
22	Male	0.8	0.8	198	20	26	6.8	3.9	1.3	27.46%
48	Female	0.8	0.2	175	48	22	8.1	4.6	1.3	42.18%

Chapter 4.Experimental Results and Discussion

The patient dataset was first divided into training data and testing data. The first 75 % of the data were used for training and remaining 25% were used for testing.

Crisp Decision Tree

The best accuracy obtained for the crisp decision tree built in Step 2 of implementation was 86.44 %. This tree is too deep with 1560 levels and it is very hard to extract the rules from. By numerous trial and error methods, the tree with best accuracy had an accuracy of 86.4 % with 1547leaves. Each rule is a path from the root node to leaf node, hence there were 1547 rules in this tree.

The system based on a crisp decision tree is quite complex to implement as there are a large number of rules. Even medical practitioners might face difficulty in understanding and evaluating these rules and the results derived from the tree.

Fuzzy Decision Tree

Fuzzy decision tree gave an accuracy of 92.38% with 37fuzzy rules. Fuzzy rules are easy to understand and the accuracy was quite high compared to the crisp decision tree. To build a fuzzy decision tree we needed fuzzy rules, considering all the factors, the fuzzy decision tree was considered over the crisp decision tree to build an inference system.

Fuzzy Inference system

Fuzzy Inference system was 88.3% accurate in being able to correctly identify the patients suffering from liver disease.

The Surface View is used to show the graphical mapping between any two inputs and any one output, the colors in the graph change according to the output values. Surface view of the output corresponds to the same input as in Table 3.As we can consider only two input values at a time, total Bilirubin and Direct Bilirubin are shown in the example surface view given in Figure 17.

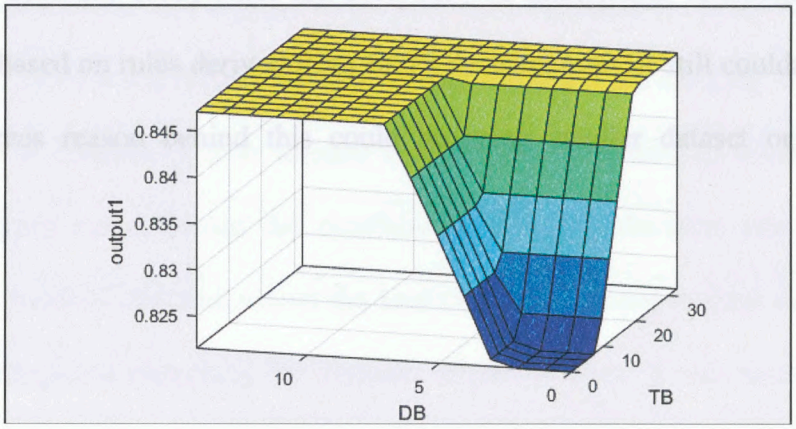


Figure 17. Surface view of FIS with DB and TB as inputs

Table 4. Comparison of Crisp DT, FDT and Inference System

	Accuracy	Number of Rules	Ease of Understanding	Sensitivity	Specificity
Crisp Decision Tree	86.67%	1547	Difficult compared to fuzzy decision tree	86.55%	86.06%
Fuzzy Decision Tree	92.38%	37	Easy to understand	97.7%	89.3%
Fuzzy Inference System	88.3%	37	Easy to understand	88.3%	87.6%

From the above Table 4 it can be observed that the fuzzy decision tree achieved better accuracy of 92.8% when compared to crisp decision tree (86.67%) and fuzzy inference system (88.3%). Reason behind fuzzy decision tree outperforming crisp decision tree could be because crisp

decision trees deal with only crisp boundary values which sometimes can be unreasonable, for example, person having body temperature of 101.9 diagnosed as low fever and person with body temperature of 102 to be diagnosed as having high fever is unreasonable. Whereas fuzzy decision trees make reasonable decisions, i.e. linguistic values are used instead of crisp values. Along with high accuracy fuzzy rules are even easy to understand. Although fuzzy inference system was built based on rules derived from fuzzy decision trees, it still couldn't outperform the fuzzy decision trees reason behind this could be using smaller dataset or the membership functions used.

Chapter 5. Conclusion and Future work

The purpose of the research described in this thesis was to build a system which could help experts or medical practitioners in preliminary diagnosing of liver disease. To achieve this goal, a crisp decision tree, fuzzy decision tree and a hybrid fuzzy rule-based inference system, which used a fuzzy decision tree to derive rules were developed. As stated in the hypotheses in Chapter 1, it is possible to build a rule-based decision support system with the help of a fuzzy decision tree that can help in diagnosing liver disease. FID software was used to build a fuzzy decision tree and its accuracy outperformed the accuracy of the crisp decision tree. As discussed in Chapter 2, fuzzy decision tree was indeed the best choice compared to crisp decision tree in the field of medical diagnosis especially for diseases related to liver. It was also observed that the artificial neural networks combined with fuzzy logic was the best choice for novice researchers who desire good results, whereas this research shows that decision trees when combined with fuzzy logic can also result in a good performance.

The rules extracted from the fuzzy decision tree were used in building the fuzzy rule based decision support system. The developed fuzzy inference system provided 87.6% accuracy in diagnosing whether the patient has liver disease or not. However, the fuzzy inference system did not outperform fuzzy decision tree in accuracy. The reason behind this could be because of using a small dataset and another reason could be the membership function that was chosen. Using different membership functions might help in obtaining better results.

Acquiring the data is the biggest challenge to perform this research. In the future, it is desirable to obtain a large dataset and apply the same methodology and compare the new results to the results obtained in this thesis. The goal is to develop a fuzzy inference system that could be used in real world.

References

1. "LiverFoundation" Retrieved 11 Nov. 2016 from http://www.liverfoundation.org/education/liverlowdown/ll1013/bigpicture/
2. Esfandiari N, Babavalian MR, Moghadam AE, Tabar V, "Knowledge discovery in medicine: Current issue and future trend." Expert Systems with Applications 41 (2014), 4434-4463.
3. Bohacik J, Kambhampati C, "Classification in a heart failure dataset with a fuzzy decision tree." Advanced Research in Scientific Areas 1 (2012), 1981-1985.
4. Hyontai S, "Improving the Prediction Accuracy of Liver Disorder Disease with Oversampling" Jan - 2012
5. Shankar V, Sugumaran V , Karthikeyan C.P , Vijayaram T.R. "Diagnosis of Hepatitis using Decision tree algorithm" July 2016
6. Suthaharan, Shan. "Decision tree learning." Machine Learning Models and Algorithms for Big Data Classification. Springer US, 2016. 237-269.
7. Caponetti, Laura, and Giovanna Castellano. "Basics of Fuzzy Logic." Fuzzy Logic for Image Processing. Springer International Publishing, 2017. 39-52.
8. Kadi I., Idri A. "Cardiovascular Dysautonomias Diagnosis Using Crisp and Fuzzy Decision Tree: A Comparative Study." May 2016
9. Kumar Y., Sahoo G. " Prediction of different types of liver diseases using rule based classification model" 2013
10. Lim CK, Yew KM, Ng KH, Abdullah BJ. "A proposed hierarchical fuzzy inference system for the diagnosis of arthritic diseases." 2002
11. Berner, Eta S. Clinical decision support systems. New York: Springer Science Business Media, LLC, 2007.

12. Mary K. "Application of Data Mining Techniques to Healthcare Data" August 2004
13. Cancer, "World Health Organization", Retrieved February 2015 from http://www.who.int/mediacentre/factsheets/fs297/en/
14. Zadeh L.A "Fuzzy sets. Information and Control" 1965
15. Nguyen Hoang Phuong, VladikKreinovich "Fuzzy Logic and its Applications in Medicine" 2000
16. Awotunde J.B., Matiluko O., Fatai O "Medical Diagnosis System Using Fuzzy Logic" June 2014
17. NonsoNnamoko, Farath Arshad, David England, JitenVora "Fuzzy Expert System for Type 2 Diabetes Mellitus (T2DM) Management Using Dual Inference Mechanism" 2013
18. Satarkar S., Ali M., "FUZZY EXPERT SYSTEM FOR THE DIAGNOSIS OF COMMON LIVER DISEASE"
19. "UCI machinery" Retrieved September 2016 from<http://archive.ics.uci.edu/ml/>
20. AACC "Lab Tests Online" Bilirubin Retrieved August 2016 from https://labtestsonline.org/understanding/analytes/bilirubin/tab/test/
21. WebMD Aspartate Aminotransferase(AGT) Test Overview Retrieved August 2016 from http://www.webmd.com/digestive-disorders/aspartate-aminotransferase-ast#1
22. WebMD Alanine Aminotransferase (ALT) Test Overview Retrieved August 2016 from http://www.webmd.com/digestive-disorders/alanine-aminotransferase-alt#1
23. Us National Laboratory of Medicine, Medline Plus "Albumin-blood (serum) test Retrieved August 2016 from https://medlineplus.gov/ency/article/003480.htm.
24. AACC "Lab Tests Online" Alkphos*Alkaline Phosphatase*Retrieved August 2016 from https://labtestsonline.org/understanding/analytes/alp/tab/glance/

25. Healthline "Total Protein test" Retrieved August 2016 from http://www.healthline.com/health/total-protein#Overview1
26. AACC "Lab Tests Online" A/G ratio Retrieved August 2016 from https://labtestsonline.org/understanding/analytes/tp/tab/test/
27. University of Stavanger, Jiaqui Ye "Using Machine Learning for Exploratory Data Analysis and Predictive Modeling" Data preprocessing.
28. Lucas Laursen11 Nov 2016 | 17:00 GMT, IEE spectrum " Doctors Still struggle to Make the Most of the Computer-Aided Diagnosis"
29. "Weka software" Retrieved September 2016 from <http://www.cs.waikato.ac.nz/ml/weka/>
30. "World Health Organization", Retrieved September 2016 from <http://www.who.int/mediacentre/factsheets/fs297/en/>
31. " The 4th Asia-Pacific Primary Liver Cancer Expert Meeting" Retrieved August 2016 from<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3881321/>
32. "Liver Cancer: Statistics <"http://www.cancer.net/cancer-types/liver-cancer/statistics>
33. Shahid, Shaouli, et al. "Factors contributing to delayed diagnosis of cancer among Aboriginal people in Australia: a qualitative study." *BMJ open* 6.6 (2016): e010909.
34. "FID: Fuzzy Decision Tree" <http://www.cs.umsl.edu/~janikow/fid/>
35. Aman Singh , Babita Pandey, "Liver disorder diagnosis using linear, nonlinear and decision tree classification algorithms",2016
36. Lin R.H, An intelligent model for liver disease diagnosis, Artif. Intell. Med. 47 (2009) 53–62. doi:10.1016/j.artmed.2009.05.005.

37. Hamamoto I, Okada S, Hashimoto T, Wakabayashi H, Maeba T, Maeta H, "Prediction of the early prognosis of the hepatectomizedpatient with hepatocellular carcinoma with a neural network.," Comput. Biol. Med. 25 (1995) 49–59.

38. Hayashi Y, Setiono R, Yoshida K, "A comparison between two neural network rule extraction techniques for the diagnosis ofhepatobiliary disorders," 20 (2000) 205–216.

39. Ozyilmaz L, Yildirim T, "Artificial neural networks for diagnosis of hepatitis disease," Proc. Int. Jt. Conf. Neural Networks, 2003. 1(2003).

40. Lee, C. C., Chung, P. C., & Chen, Y. J. (2005, May). "Classification of liver diseases from ct images using bp-cmac neural network. In Cellular Neural Networks and Their Applications," 2005 9th International Workshop on (pp. 118-121). IEEE.

41. Yahagi T, "Ultrasonographic classification of Cirrhosis based on pyramid neural network," in: Can. Conf. Electr. Comput. Eng. 2005.,IEEE, 2005: pp. 1682–1685.

42. Azaid S.A., Fakhr M.W., Mohamed F, "Automatic Diagnosis of Liver Diseases from Ultrasound Images," 2006 Int. Conf. Comput.Eng. Syst. (2006) 313–319.

43. Badawi, Ahmed M., Ahmed S. Derbala, and Abou-Bakr M. Youssef. "Fuzzy logic algorithm for quantitative tissue characterization of diffuse liver diseases from ultrasound images." International Journal of Medical Informatics 55.2 (1999): 135-147.

44. Gadaras, Ioannis, and Ludmil Mikhailov. "An interpretable fuzzy rule-based classification methodology for medical diagnosis." Artificial Intelligence in Medicine 47.1 (2009): 25-41.

45. Luukka P, "Fuzzy beans in classification." Expert Syst. Appl. 38 (2011) 4798–4801.

46. Ming Lim Kian, Loo Chu Kiong, and Lim Way Soong. "Autonomous and deterministic supervised fuzzy clustering with data imputation capabilities." Applied Soft Computing 11.1 (2011): 1117-1125.

47. Neshat, M., et al. "Fuzzy expert system design for diagnosis of liver disorders." Knowledge Acquisition and Modeling, 2008. KAM'08. International Symposium on. IEEE, 2008.

48. Singh, Aman, and Babita Pandey. "Intelligent techniques and applications in liver disorders: a survey." International Journal of Biomedical Engineering and Technology 16.1 (2014): 27-70.

49. Kaur, Parminder, and Aditya Khamparia. "CLASSIFICATION OF LIVER BASED DISEASES USING RANDOM TREE." International Journal of Advances in Engineering & Technology 8.3 (2015): 306.

50. Hüllermeier, Eyke, and Stijn Vanderlooy. "Why fuzzy decision trees are good rankers." IEEE Transactions on Fuzzy Systems 17.6 (2009): 1233-1244.

Appendix: Fuzzy rules extracted from the fuzzy decision tree

1. *If (AGE is Old) and (TB is Normal) and (DB is Normal) and (SGOT is Normal) and (TP is Low) and (AG is Low) then (output1 is High)*
2. *If (AGE is Old) and (TB is Normal) and (DB is Normal) and (SGOT is Normal) and (TP is Low) and (AG is Normal) then (output1 is High).*
3. *If (AGE is Young) and (TB is Normal) and (DB is Normal) and (SGPT is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) then (output1 is low).*
4. *If (AGE is Young) and (TB is Normal) and (DB is Normal) and (SGPT is High) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) then (output1 is low).*
5. *If (AGE is Adult) and (DB is Normal) and (AAP is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Low) then (output1 is High) .*
6. *If (AGE is Adult) and (DB is Normal) and (AAP is High) and (SGOT is Normal) and (TP is Normal) and (ALB is Low) then (output1 is low) .*
7. *If (AGE is Adult) and (GENDER is Male) and (TB is Normal) and (DB is Normal) and (AAP is Normal) and (SGPT is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) then (output1 is low) .*
8. *If (AGE is Adult) and (GENDER is Male) and (TB is Normal) and (DB is Normal) and (AAP is High) and (SGPT is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) then (output1 is low) .*
9. *If (AGE is Adult) and (GENDER is Male) and (DB is Normal) and (SGPT is High) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) then (output1 is low)*

.

10. *If (AGE is Adult) and (GENDER is Female) and (TB is Normal) and (DB is Normal) and (AAP is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) and (AG is Low) then (output1 is High) .*
11. *If (AGE is Adult) and (GENDER is Female) and (TB is Normal) and (DB is Normal) and (AAP is High) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) and (AG is Low) then (output1 is High) .*
12. *If (AGE is Adult) and (GENDER is Female) and (TB is Normal) and (DB is Normal) and (AAP is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) and (AG is Normal) then (output1 is High) .*
13. *If (AGE is Adult) and (GENDER is Female) and (TB is Normal) and (DB is Normal) and (AAP is High) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) and (AG is Normal) then (output1 is High) .*
14. *If (AGE is Adult) and (DB is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is High) then (output1 is low) .*
15. *If (AGE is Old) and (GENDER is Male) and (TB is Normal) and (DB is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Low) and (AG is Low) then (output1 is High) .*
16. *If (AGE is Old) and (GENDER is Male) and (TB is Normal) and (DB is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) and (AG is Low) then (output1 is High) .*
17. *If (AGE is Old) and (GENDER is Male) and (TB is Normal) and (DB is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is High) and (AG is Low) then (output1 is low) .*

18. *If (AGE is Old) and (GENDER is Male) and (TB is Normal) and (DB is Normal) and (AAP is Normal) and (SGOT is Normal) and (TP is Normal) and (AG is Low) then (output1 is High) .*
19. *If (AGE is Old) and (GENDER is Male) and (TB is Normal) and (DB is Normal) and (AAP is High) and (SGOT is Normal) and (TP is Normal) and (AG is Low) then (output1 is High) .*
20. *If (AGE is Old) and (TB is Normal) and (DB is Normal) and (AAP is Normal) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) and (AG is Normal) then (output1 is High) .*
21. *If (AGE is Old) and (TB is Normal) and (DB is Normal) and (AAP is High) and (SGOT is Normal) and (TP is Normal) and (ALB is Normal) and (AG is Normal) then (output1 is High) .*
22. *If (AGE is Young) and (DB is Normal) and (SGOT is Normal) and (TP is High) then (output1 is High) .*
23. *If (AGE is Adult) and (DB is Normal) and (SGPT is Normal) and (SGOT is Normal) and (TP is High) and (AG is Low) then (output1 is High) .*
24. *If (AGE is Adult) and (DB is Normal) and (SGPT is High) and (SGOT is Normal) and (TP is High) then (output1 is High) .*
25. *If (AGE is Old) and (DB is Normal) and (SGOT is Normal) and (TP is High) then (output1 is low) .*
26. *If (TB is Normal) and (DB is Normal) and (SGPT is Normal) and (SGOT is High) and (AG is Low) then (output1 is High) .*

27. *If (TB is Normal) and (DB is Normal) and (SGPT is Normal) and (SGOT is High) and (AG is Normal) then (output1 is low) .*
28. *If (GENDER is Male) and (DB is Normal) and (SGPT is High) and (SGOT is High) then (output1 is High).*
29. *If (GENDER is Female) and (DB is Normal) and (SGPT is High) and (SGOT is High) then (output1 is High).*
30. *If (TB is Normal) and (DB is High) and (AAP is Normal) then (output1 is High .*
31. *If (TB is Normal) and (DB is High) and (AAP is High) then (output1 is High) .*
32. *If (TB is High) and (DB is High) then (output1 is High) .*
33. *If (DB is Normal) and (SGPT is Normal) and (SGOT is Normal) and (TP is High) and (AG is Normal) then (output1 is High) .*

I have submitted this thesis in partial fulfillment of the requirements for the degree of Master of Science

8/16/17 — Date

Himaja — Himaja Sivaraju

We approve the thesis of Himaja Sivaraju as presented here.

7/17/17 — Date

Shamim Khan, Professor of Computer Science. Thesis Advisor

7/17/17 — Date

Rania Hodhod. Assistant Professor of Computer Science

7/14/17 — Date

Sumanth Yenduri, Associate Professor of Computer Science

8/9/17 — Date

Ronald Linton, Professor of Mathematics

7/17/17 — Date

Wayne Summers, Professor and Chairperson, TSYS School of Computer Science